CSE574 Introduction to Machine Learning

Neural Network: Forward Pass and Back-propagation

Jue Guo

University at Buffalo

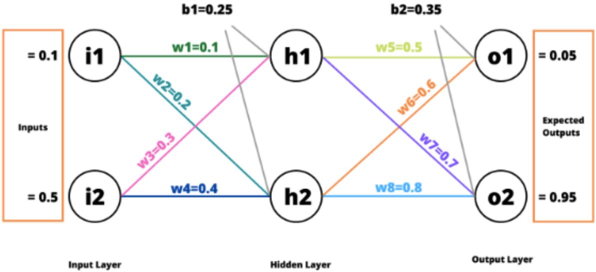
March 1, 2024

Outline

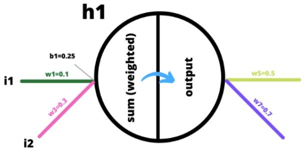
Introduction

- step by step forward pass (forward propagation) and backward pass (back-propagation)
- a single hidden layer neural network
    - solving one complete cycle of forward and back-propagation

A simple neural network:



Peeking inside a single neuron



Inside a unit, two operations happen

1. computation of weighted sum and
2. squashing of the weighted sum using an activation function.

The result from the activation function (Sigmoid function (Logistic function) as the activation function.) becomes an input to the next layer (until the next layer is an Output Layer).

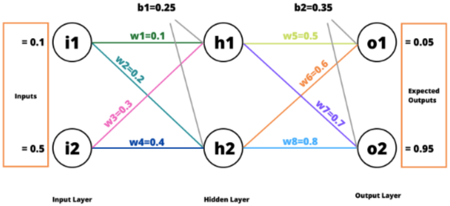
Forward Pass

For h_1

$$\mathsf{sum}_{h1} = i_1 * w_1 + i_2 * w_3 + b_1$$
$$\mathsf{sum}_{h1} = 0.1 * 0.1 + 0.5 * 0.3 + 0.25 = 0.41$$

Now we pass this weighted sum through the logistic function (sigmoid function) so as to squash the weighted sum into the range (0 and +1). The logistic function is an activation function for our example neural network.

What is the output?

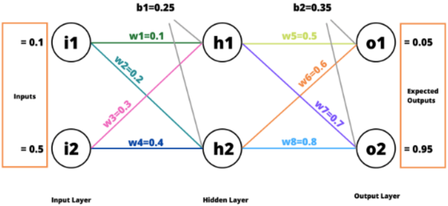


$$\text{output}_{h1} = \frac{1}{1+e^{-\text{sum}_{h1}}}$$

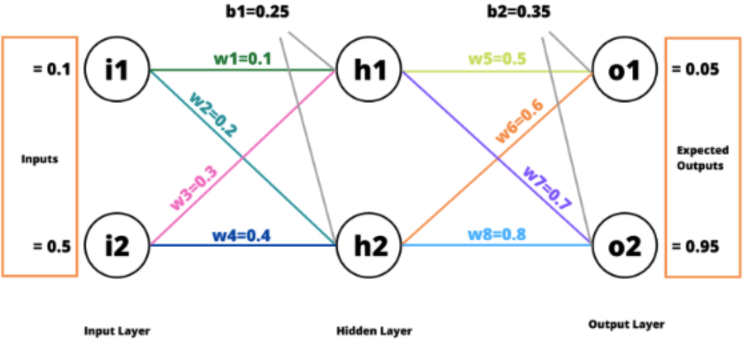
$$\text{output}_{h1} = \frac{1}{1+e^{-0.41}} = 0.60108$$

Similarly for h2, we perform the weighted sum operation sum_{h2} and compute the activation value output $_{h2}$.

calculate...



$$\text{sum}_{h2} = i_1 * w_2 + i_2 * w_4 + b_1 = 0.47$$
$$\text{output}_{h2} = \frac{1}{1+e^{-\text{sum}_{h2}}} = 0.61538$$



Now do it for the output ...

For o1,

$$\text{sum}_{o1} = \text{output}_{h1} * w_5 + \text{output}_{h2} * w_6 + b_2 = 1.01977$$

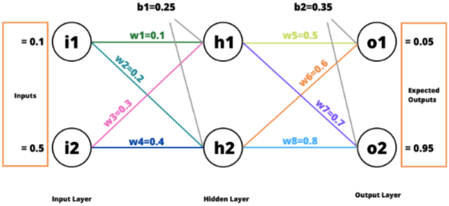
$$\text{output}_{o1} = \frac{1}{1+e^{-\text{sum}_{o1}}} = 0.73492$$

Similarly for o2,

$$\text{sum}_{o2} = \text{output}_{h1} * w_7 + \text{output}_{h2} * w_8 + b_2 = 1.26306$$

$$\text{output}_{o2} = \frac{1}{1+e^{-\text{sum}_{o2}}} = 0.77955$$

Computing the Total Error



We started off supposing the expected outputs to be 0.05 and 0.95 respectively for output $_{o1}$ and output $_{o2}$.

- Now we will compute the errors based on the outputs received until now and the expected outputs.
- Use Mean Squared Error, **Calculate** ...

We'll use the following error formula,

$$E_{\text{total}} = \sum \frac{1}{2}(\text{ target } - \text{ output })^2$$

To compute E_{total}, we need to first find out respective errors at $o1$ and $o2$

$$E_1 = \frac{1}{2}(\text{ target }_1 - \text{ output }_{o1})^2$$

$$E_1 = \frac{1}{2}(0.05 - 0.73492)^2 = 0.23456$$

Similarly for E2,

$$E_2 = \frac{1}{2}(\text{ target }_2 - \text{ output }_{o2})^2$$

$$E_2 = \frac{1}{2}(0.95 - 0.77955)^2 = 0.01452$$

Therefore,

$$E_{\text{total}} = E_1 + E_2 = 0.24908$$

The Back-propagation

The aim of backpropagation (backward pass) is to distribute the total error back to the network so as to update the weights in order to minimize the cost function (loss).

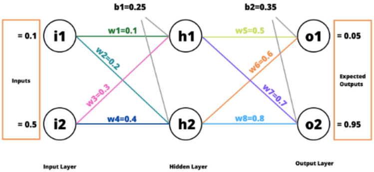
- The weights are updated in such as way that when the next forward pass utilizes the updated weights, the total error will be reduced by a certain margin (until the minima is reached).

Chain Rule in Calculus

If we have $y = f(u)$ and $u = g(x)$ then we can write the derivative of y as:

$$\frac{dy}{dx} = \frac{dy}{du} * \frac{du}{dx}$$

w5, w6, w7, w8



For weights in the output layer (w5, w6, w7, w8); Let's first see w5. If we become clear on how w5 is updated, then it would be really easy for us to generalize the same to the rest of the weights.

If we look closely at the example neural network,

- E_1 is affected by output_{o1}
- output_{o1} is affected by sum_{o1}
- sum_{o1} is affected by $w5$

$$\frac{\partial E_{\text{total}}}{\partial w5} = \frac{\partial E_{\text{total}}}{\partial output_{o1}} * \frac{\partial output_{o1}}{\partial sum_{o1}} * \frac{\partial sum_{o1}}{\partial w5}$$

Component 1: partial derivative of Error w.r.t. Output

$$E_{\text{total}} = \sum \tfrac{1}{2}(\text{ target } - \text{ output })^2$$

$$E_{\text{total}} = \tfrac{1}{2}(\text{ target }_1 - \text{ output }_{o1})^2 + \tfrac{1}{2}(\text{ target }_2 - \text{ output }_{o2})^2$$

Therefore,

$$\begin{aligned}\frac{\partial E_{\text{total}}}{\partial \text{ output }_{o1}} &= 2 * \frac{1}{2} * (\text{ target }_1 - \text{ output }_{o1}) * -1 \\ &= \text{ output }_{o1} - \text{ target }_1\end{aligned}$$

Component 2: partial derivative of Output w.r.t. Sum

The output section of a unit of a neural network uses non-linear activation functions. The activation function used in this example is Logistic Function. When we compute the derivative of the Logistic Function, we get:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\frac{\mathrm{d}}{\mathrm{dx}}\sigma(x) = \sigma(x)(1 - \sigma(x))$$

Therefore, the derivative of the Logistic function is equal to output multiplied by (1 – output).

$$\frac{\partial \text{ output }_{o1}}{\partial \text{ sum }_{o1}} = \text{ output }_{o1} (1 - \text{ output }_{o1})$$

Component 3: partial derivative of Sum w.r.t. Weight

$$\text{sum}_{o1} = \text{output}_{h1} * w_5 + \text{output}_{h2} * w_6 + b_2$$

Therefore,

$$\frac{\partial sum_{o1}}{\partial w5} = \text{output}_{h1}$$

Putting them together,

$$\frac{\partial E_{\text{total}}}{\partial w5} = \frac{\partial E_{\text{total}}}{\partial \text{ output}_{o1}} * \frac{\partial \text{ output}_{o1}}{\partial sum_{o1}} * \frac{\partial sum_{o1}}{\partial w5}$$

$$\frac{\partial E_{total}}{\partial w5} = [\text{output}_{o1} - \text{target}_1] * [\text{output}_{o1}(1 - \text{output}_{o1})] * [\text{output}_{h1}]$$

$$\frac{\partial E_{total}}{\partial w5} = 0.68492 * 0.19480 * 0.60108$$

$$\frac{\partial E_{total}}{\partial w5} = 0.08020$$

The new_w_5 is,

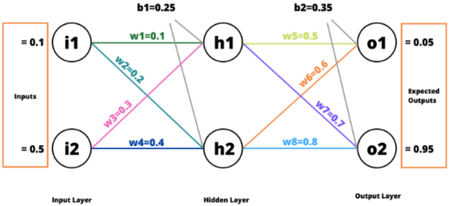
$new\_w_5 = w5 - n * \frac{\partial E_{\text{total}}}{\partial w5}$, where n is learning rate.

$$new\_w_5 = 0.5 - 0.6 * 0.08020$$

$$new\_w_5 = 0.45187$$

We can proceed similarly for w6, w7 and w8. (10 minutes)

w1,w2,w3,w4



$$\frac{\partial E_1}{\partial w1} = \frac{\partial E_1}{\partial output_{o1}} * \frac{\partial output_{o1}}{\partial sum_{o1}} * \frac{\partial sum_{o1}}{\partial output_{h1}} * \frac{\partial output_{h1}}{\partial sum_{h1}} * \frac{\partial sum_{h1}}{\partial w1}$$

$$\frac{\partial E_2}{\partial w1} = \frac{\partial E_2}{\partial output_{o2}} * \frac{\partial output_{o2}}{\partial sum_{o2}} * \frac{\partial sum_{o2}}{\partial output_{h1}} * \frac{\partial output_{h1}}{\partial sum_{h1}} * \frac{\partial sum_{h1}}{\partial w1}$$